



# School of Full Stack

Data Visualization with R

# Contents

1. Why we need Data Viz?
2. Exploratory Data Analysis
3. Basic Plots (base R)
4. Introduction to ggplot2
  - a. Histogram
  - b. Density
  - c. Barplot
  - d. Boxplot
  - e. Scatter Plot
  - f. Smooth
  - g. Labels
  - h. Theme + Facet + Color Scale

# Why we need data viz?

Can you tell the relationship between these variables?

mpg

Copy

```
#> # A tibble: 234 x 11
```

```
#>   manufacturer model displ  year   cyl trans      drv   cty   hwy fl   class
```

```
#>   <chr>         <chr> <dbl> <int> <int> <chr>   <chr> <int> <int> <chr> <chr>
```

```
#> 1 audi         a4      1.8  1999     4 auto(l5) f       18    29 p    comp
```

```
#> 2 audi         a4      1.8  1999     4 manual(m5) f       21    29 p    comp
```

```
#> 3 audi         a4      2    2008     4 manual(m6) f       20    31 p    comp
```

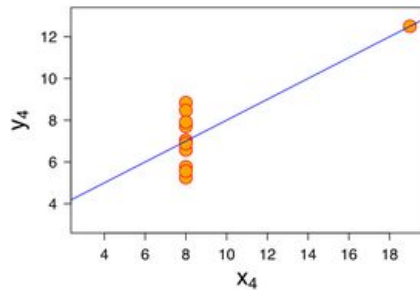
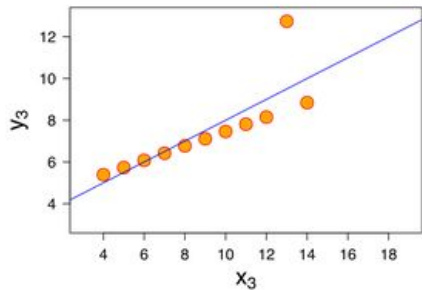
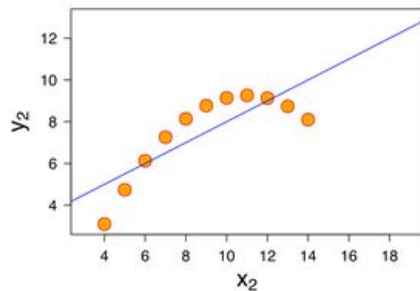
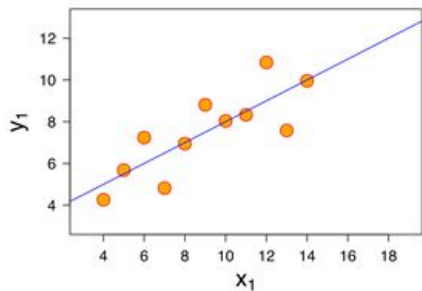
```
#> 4 audi         a4      2    2008     4 auto(av) f       21    30 p    comp
```

```
#> 5 audi         a4      2.8  1999     6 auto(l5) f       16    26 p    comp
```

```
#> 6 audi         a4      2.8  1999     6 manual(m5) f       18    26 p    comp
```

```
#> # ... with 228 more rows
```

# Why we need data viz?



All 4 datasets have the same statistics e.g.

- mean
- sd
- correlation

[What does it mean if two data sets have the same standard deviation? - Quora](#)

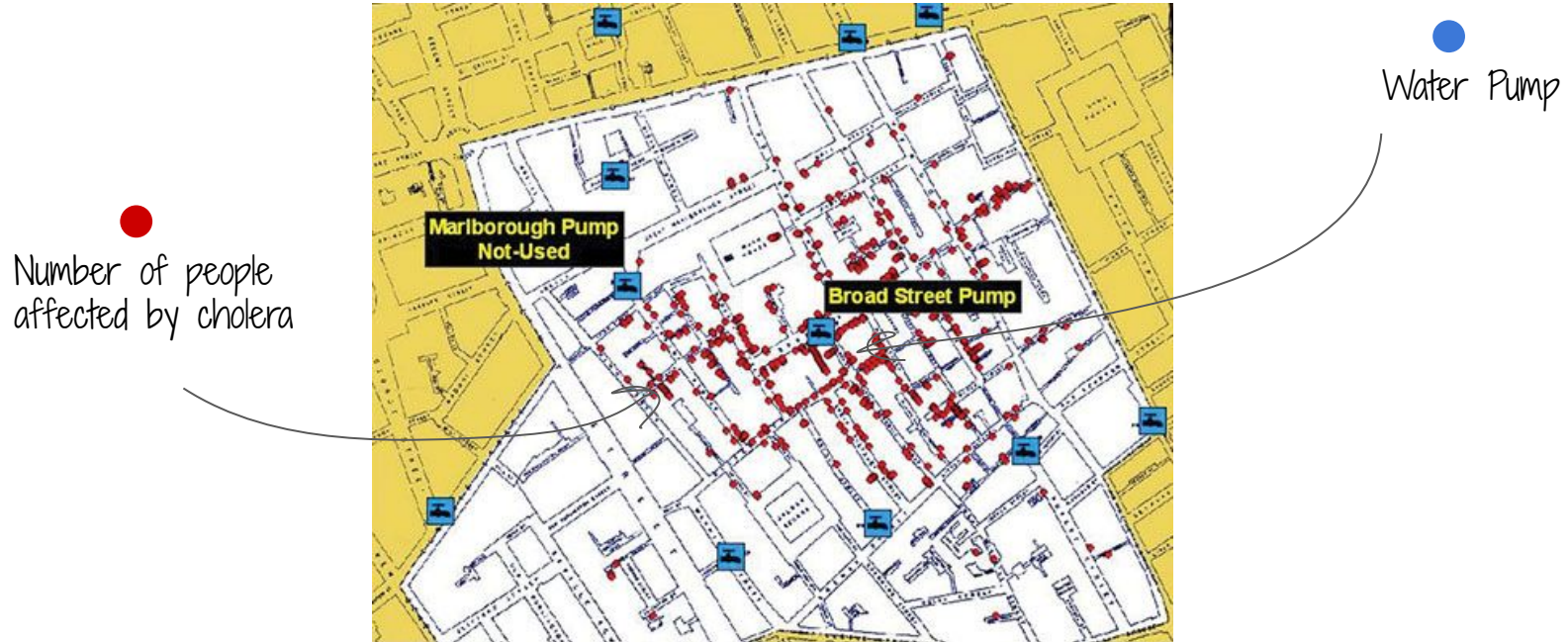
# Great Benefits of Data Viz



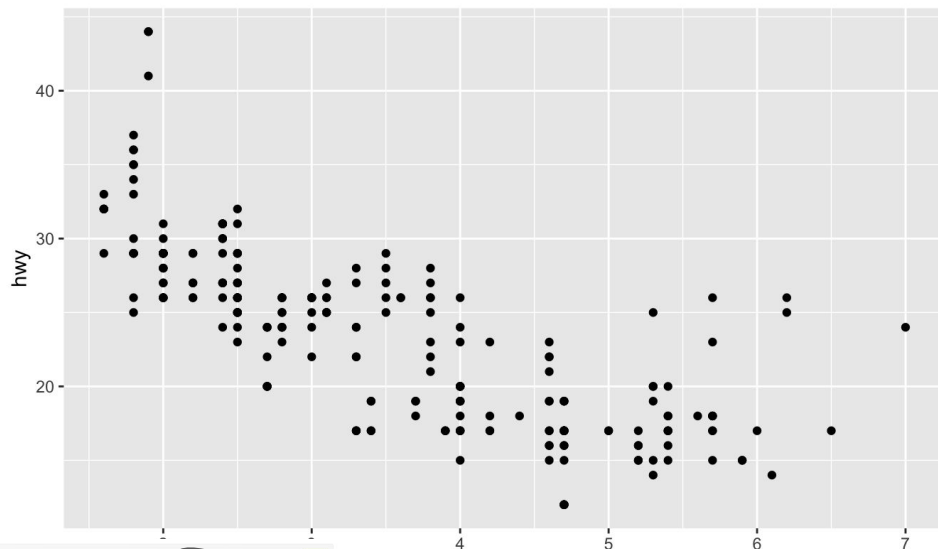
John Tukey

“The greatest value of a picture is when it forces us to notice what **we never expected to see**”

# Case Study - John Snow



# Exploratory Data Analysis



EDA can be done in two ways

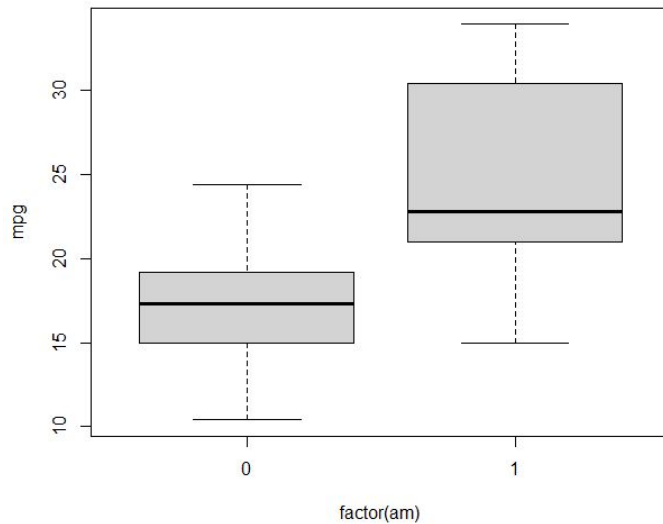
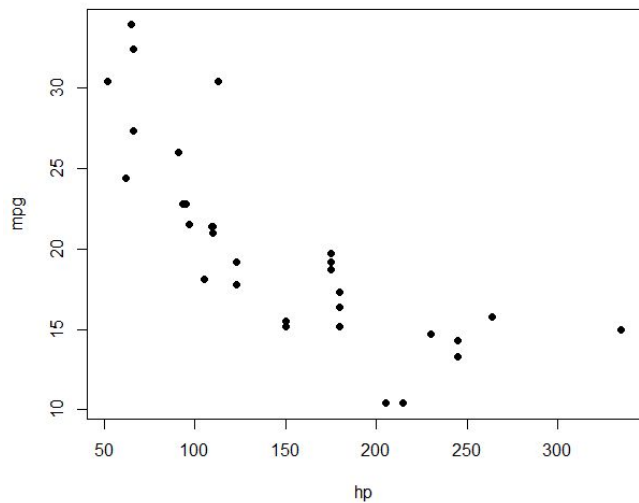
1. Numerical Method
  - a. summary stats
  - b. basic modeling
2. Graphical Method

```
mpg
#> # A tibble: 234 x 11
#>   manufacturer model displ  year  cyl trans      drv  cty   hwy fl   class
#>   <chr>          <chr>  <dbl> <int>   <int> <chr>   <chr> <int> <int> <chr> <chr>
#> 1 audi         a4           1.8  1999     4 auto(l5) f      18   29 p   com
#> 2 audi         a4           1.8  1999     4 manual(m5) f      21   29 p   com
#> 3 audi         a4           2.0  2008     4 manual(m5) f      20   31 p   com
#> 4 audi         a4           2.0  2008     4 auto(su) f      21   30 p   com
#> 5 audi         a4           2.8  1999     6 auto(l5) f      16   26 p   com
#> 6 audi         a4           2.8  1999     6 manual(m5) f      18   26 p   com
#> # ... with 228 more rows
```

displ

We turn raw data into chart

# Basic Plots in R





# Intro to ggplot2



# Our Template

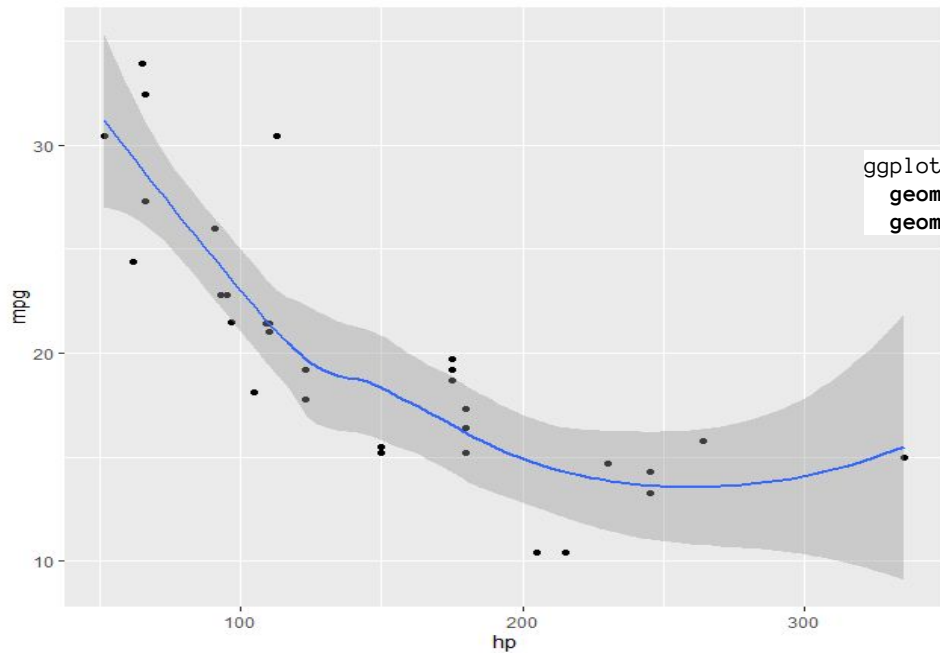


This can generate more  
than 30 charts



```
ggplot(data = ... , mapping = aes( ... ) ) +  
  geom_...()
```

# Our Very First Plot

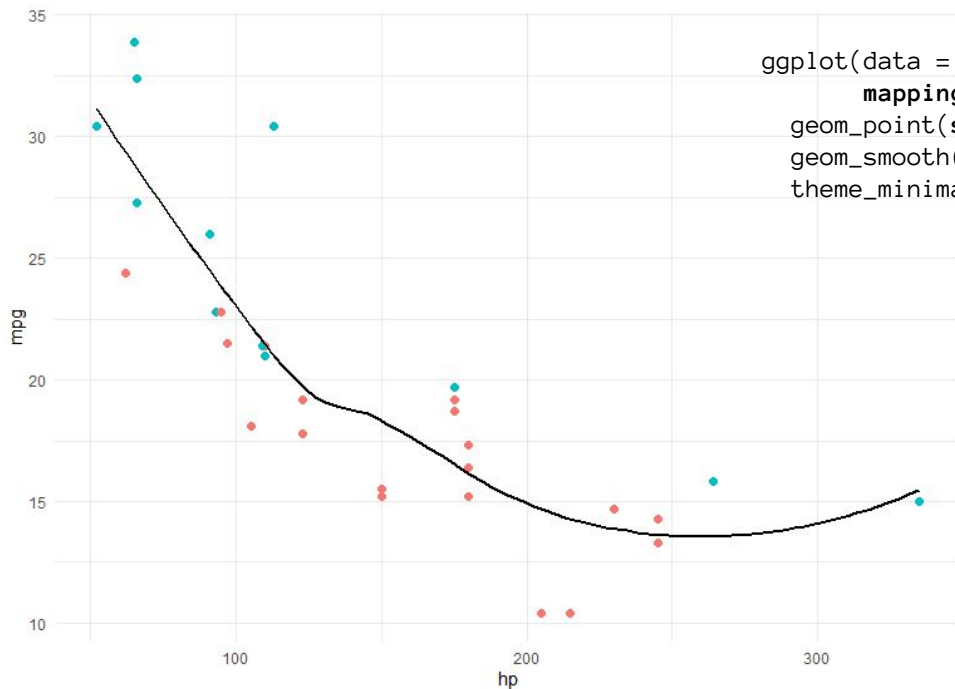


```
ggplot(data = mtcars, mapping = aes(x = hp, y = mpg)) +  
  geom_point() +  
  geom_smooth()
```

Practice first, explain later



# Setting vs. Mapping



```
ggplot(data = mtcars,  
  mapping = aes(x = hp, y = mpg, col = factor(am))) +  
  geom_point(size = 2) +  
  geom_smooth(col = "black", se = FALSE) +  
  theme_minimal()
```

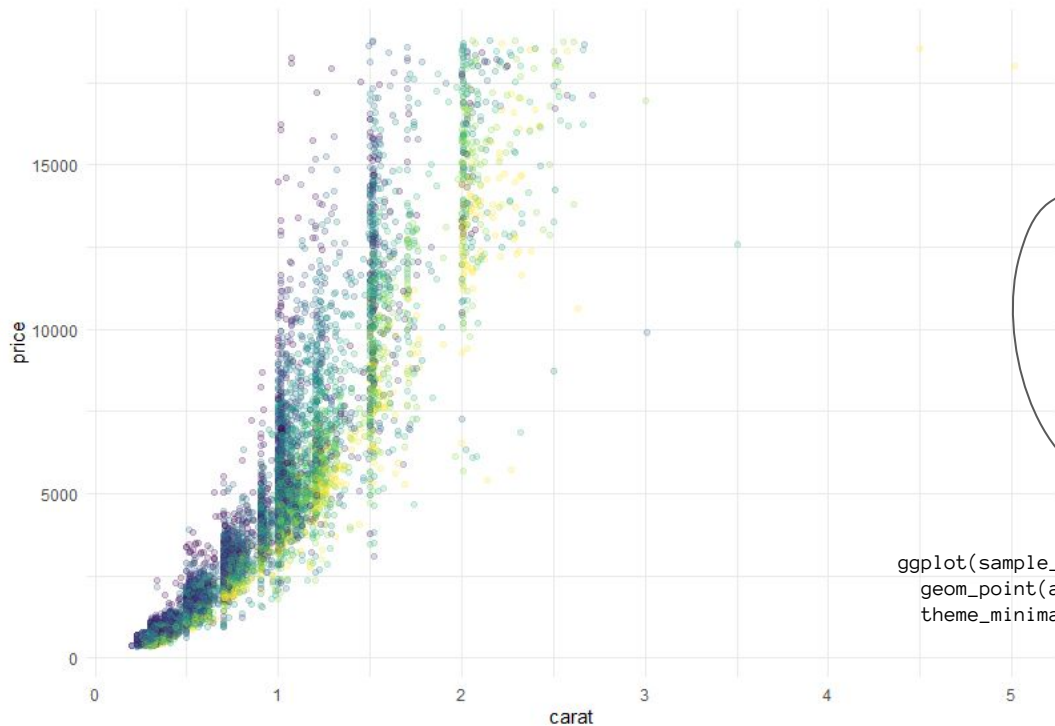
Mapping

Inside aes()

Setting

Outside aes()

# We Can Map Variable to Color



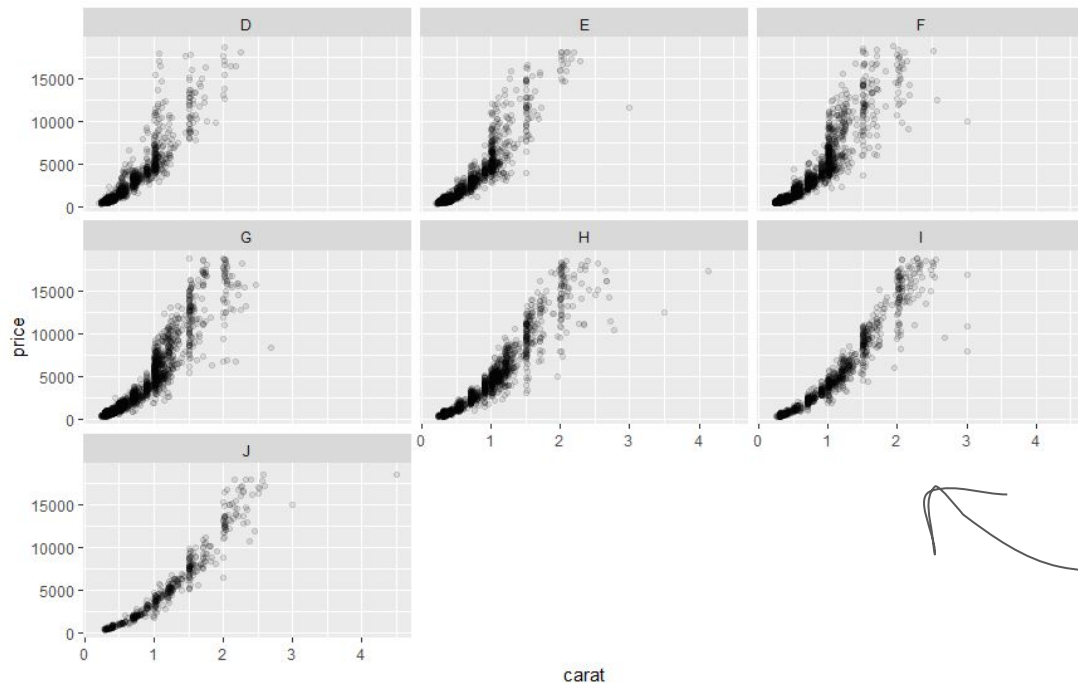
Mapping color variable to color  
aesthetic inside aes()

color

- D
- E
- F
- G
- H
- I
- J

```
ggplot(sample_n(diamonds, 10000), aes(carat, price, col=color)) +  
  geom_point(alpha=0.2) +  
  theme_minimal()
```

# Facet



Small Multiples



# School of Full Stack

<https://datarockie.com>